8:30 — Guillaume Belanger

*The most sensitive period search we can do on event data*

*Treating detection of transient phenomena as a statistical problem*

9:00 — Vinay Kashyap

*Flux measurement from high-resolution spectra*

9:10 — Discussion: *A Practical Guide to Data Analysis*

9:25 — Discussion: modeling backgrounds

9:30 — Herman Marshall

*Concordance*
Flux Measurement from High-Resolution Spectra

Vinay Kashyap
CXC/CfA
From Photons to Flux

Vinay Kashyap
CXC/CfA
If you add up the fluxes for all photons

- The obvious thing
  - Compute erg/cm² for each bin, add them up, divide by exposure time.
  - Works fine!

- But what if you have background? Observed counts are a mixture of true source events and background events (and you don’t know which is which)
  - Again do the obvious thing — estimate what fraction of each count could be background and remove that fraction before estimating the flux
  - Does not work!
If you add up the fluxes
Dealing with background

❖ How about if we can figure out the probability that a given photon is from the background?
❖ use Monte Carlo to randomly generate background events and superpose on observed spectrum
❖ pick off the nearest event as background
❖ repeat 100s of times, keep track of how many times a given observed count gets flagged — the more often, the higher the probability that it is a background event
example
Dealing with background

- Dealing with background
  - $\sigma$ mean
  - true $\pm 1\sigma$

![Graph with N_{src}=21.0, logT=6.80 showing flux estimated/true vs N_{src} with 1σ and 2σ bounds, mean, and mode.](image-url)
tl;dl

- If you add up photon energies when you know the energies accurately, background contamination will cause a large bias.
- There may be several mitigating strategies, but one that does work is to do Monte Carlo sampling based on assigned/computed probability that a given event comes from the background.
A Practical Guide to Analysis

Vinay Kashyap
CXC/CfA
Action item from IACHEC 2018

Write a practical guide to best practices in X-ray analysis

Need co-authors!

And feedback!
A straw man ToC

1. What is this about
   1. An opinionated document
   2. listing things useful to cal scientists, and only things useful to cal scientists
2. Jargon and notation
   1. a dictionary of stats terms
3. On error bars
   1. there is a zoo of them (frequentist vs Bayesian, 1σ vs 90%, equal-tail vs highest posterior density)
   2. how to propagate, when propagation doesn’t work
4. Sampling
   1. about pseudorandom numbers
   2. Distributions (are not functions)
   3. Monte Carlo and Bootstrap (types, what to be careful about)
   4. Tips and tricks
A straw man ToC (contd.)

5. Likelihood and Probability
   1. Conditional probability, Bayes’ Theorem

6. Model Fitting
   1. Models, parameters, estimates, uncertainty intervals
   2. Non-parametric analysis (Kernel-density estimates, smoothing, histograms, Bayesian Blocks)
   3. log-Likelihood, chisq, Cash/cstat, Wstat — compare and contrast
   4. optimization techniques, MCMC

7. p-values and Hypothesis Tests
   1. What does a p-value mean and what it doesn’t mean
   2. K-S test, F-test, Bayes Factors
   3. Examples: source detection

8. New Directions
   1. ML (Neural Networks, Gaussian Processes)
8:30 — Guillaume Belanger

*The most sensitive period search we can do on event data*
*Treating detection of transient phenomena as a statistical problem*

9:00 — Vinay Kashyap

*Flux measurement from high-resolution spectra*

9:10 — Discussion: *A Practical Guide to Data Analysis*

9:25 — Discussion: modeling backgrounds

9:30 — Herman Marshall

*Concordance*